

Easy Estimation of Normalizing Constants and Bayes Factors from Posterior Simulation: Stabilizing the Harmonic Mean Estimator ¹

Jaya M. Satagopan
Memorial Sloan-Kettering Cancer Center

Michael A. Newton
University of Wisconsin

Adrian E. Raftery
University of Washington

Technical Report no. 382
Department of Statistics
University of Washington

November 2000

¹Jaya M. Satagopan is Assistant Professor, Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, NY 10021; Email: satago@biosta.mskcc.org. Michael A. Newton is Associate Professor of Statistics and Biostatistics, University of Wisconsin, Madison, WI 53706, Email: newton@biostat.wisc.edu, Web: www.stat.wisc.edu/~newton. Adrian E. Raftery is Professor of Statistics and Sociology, Department of Statistics, University of Washington, Box 354322, Seattle WA 98195-4322; Email: raftery@stat.washington.edu; Web: www.stat.washington.edu/raftery. Raftery's research was supported by ONR grant no. N00014-96-1-1092.

Report Documentation Page				Form Approved OMB No. 0704-0188	
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE NOV 2000		2. REPORT TYPE		3. DATES COVERED 00-11-2000 to 00-11-2000	
4. TITLE AND SUBTITLE Easy Estimation of Normalizing Constants and Bayes Factors from Posterior Simulation: Stabilizing the Harmonic Mean Estimator				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) University of Washington, Department of Statistics, Box 354322, Seattle, WA, 98195-4322				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES 22	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

Abstract

The Bayes factor is a useful summary for model selection. Calculation of this measure involves evaluating the integrated likelihood (or prior predictive density), which can be estimated from the output of MCMC and other posterior simulation methods using the harmonic mean estimator. While this is a simulation-consistent estimator, it can have infinite variance. In this article we describe a method to stabilize the harmonic mean estimator. Under this approach, the parameter space is reduced such that the modified estimator involves a harmonic mean of heavier tailed densities, thus resulting in a finite variance estimator. We discuss general conditions under which this reduction is applicable and illustrate the proposed method through several examples.

Keywords: Bayes factor, Beta-binomial, Integrated likelihood, Poisson-Gamma distribution, Statistical genetics, Variance reduction.

Contents

1	Introduction	1
2	Stabilizing the Harmonic Mean Estimator	2
3	Statistical Genetics	6
4	Beta-Binomial	8
5	Other Reductions: A Simple Poisson-Gamma Model	14
6	Discussion	14

List of Figures

1	Normal and t -based Harmonic Mean Estimates	4
2	Variability of Different Estimators of the Integrated Likelihood	5
3	Bayes Factors for the Flowering Time Data	9
4	Variability of the Bayes Factor Estimates for the Flowering Time Data . . .	10
5	Comparison of Estimators of the Integrated Likelihood for the NBA Data . .	13

1 Introduction

The integrated likelihood is an important quantity in model comparison — it is the key component of the Bayes factor for example (Kass and Raftery 1995). Consider data y , a likelihood function $\pi(y|\theta)$ from a model for y indexed by a parameter θ , in which both y and θ may be vector-valued, and a prior distribution $\pi(\theta)$. The integrated likelihood of y is then defined as

$$\pi(y) = \int \pi(y|\theta)\pi(\theta) d\theta.$$

The integrated likelihood normalizes the product of the likelihood and the prior in forming the posterior density $\pi(\theta|y)$. Furthermore, as a function of y prior to data collection, $\pi(y)$ is the prior predictive density.

Evaluating the integrated likelihood can present a difficult computational problem. Newton and Raftery (1994) showed that $\pi(y)$ can be expressed as an expectation with respect to the posterior distribution of the parameter, thus motivating an estimate based on a Monte Carlo sample from the posterior. By Bayes's theorem,

$$\frac{1}{\pi(y)} = \int \frac{\pi(\theta|y)}{\pi(y|\theta)} d\theta = E \left\{ \frac{1}{\pi(y|\theta)} \middle| y \right\}, \quad (1)$$

suggesting that the integrated likelihood $\pi(y)$ can be approximated by the harmonic mean

$$\hat{\pi}_{\text{HM}}(y) = \left[\frac{1}{B} \sum_{t=1}^B \frac{1}{\pi(y|\theta^t)} \right]^{-1}, \quad (2)$$

based on B draws $\theta^1, \theta^2, \dots, \theta^B$ from the posterior distribution $\pi(\theta|y)$. This sample might come out of a standard Markov chain Monte Carlo implementation, for example. Though $\hat{\pi}_{\text{HM}}(y)$ is consistent as the simulation size B increases, its precision is not guaranteed.

The simplicity of the harmonic mean estimator (2) is its main advantage over other more specialized techniques (e.g. Chib 1995, Raftery 1996, Lewis and Raftery 1997, DiCiccio et al. 1997). It uses only within-model posterior samples and likelihood evaluations which are often available anyway as part of posterior sampling. A major drawback of the harmonic mean estimator is its computational instability. The estimator is consistent but may have infinite variance (measured by $\text{Var}\{[\pi(y|\theta)]^{-1}|y\}$) across simulations, even in simple models. When this is the case, one consequence is that when the cumulative estimate of the harmonic mean estimate (2) based on the first B draws from the posterior is plotted against B , the plot has occasional very large jumps, and looks unstable. In this article we present a method to stabilize the harmonic mean estimator. We develop general conditions under which this method works and we demonstrate the method in several examples.

2 Stabilizing the Harmonic Mean Estimator

An overly simple but helpful example to illustrate our proposed method is the model in which $\theta = (\mu, \psi)$ records the mean and precision of a single normally distributed data point y . A conjugate prior is $\psi \sim \text{Gamma}(\alpha/2, \alpha/2)$, and

$$(\mu|\psi) \sim \text{Normal}(\mu_0, n_0\psi)$$

where α, n_0 , and μ_0 are hyperparameters (e.g., Bernardo and Smith, 1994, page 268 or Appendix I). The integrated likelihood $\pi(y)$ is readily determined to be the ordinate of a t density, $\text{St}(y|\mu_0, n_0/(n_0 + 1), \alpha)$ in the notation of Bernardo and Smith (1994, page 122 or Appendix I). Were we to approximate $\pi(y)$ using equation (2), instead of taking the analytically determined value, we could measure the stability of the estimator with the variance $\text{Var}\{[\pi(y|\theta)]^{-1}|y\}$. This variance, in turn, is determined by the second noncentral moment $\text{E}\{[\pi(y|\theta)]^{-2}|y\}$ which is proportional to

$$\int \int \psi^{\alpha/2} \exp \left\{ \frac{\psi}{2} [(y - \mu)^2 - n_0(\mu - \mu_0)^2 - \alpha] \right\} d\psi d\mu,$$

and which is infinite in this example owing to the divergence of the integral in μ for each ψ . The reciprocal of the light-tailed normal density forms too large an integrand to yield a finite posterior variance, and hence the harmonic mean estimator is unstable.

An alternative estimator, supported equally by the basic equation (1), is

$$\hat{\pi}_{\text{SHM}}(y) = \left[\frac{1}{B} \sum_{t=1}^B \frac{1}{\pi(y|\mu^t)} \right]^{-1}, \quad (3)$$

which we call a stabilized harmonic mean. In (3), μ^t is the mean component of $\theta^t = (\mu^t, \psi^t)$, and thus is a draw from the marginal posterior distribution $\pi(\mu|y)$. The stabilized harmonic mean is formed not from standard likelihood values, but rather from marginal likelihoods obtained by integrating out the precision parameter ψ . It is straightforward to show that this marginal likelihood has the form of a t ordinate,

$$\pi(y|\mu) = \text{St} \left\{ y|\mu, (\alpha + 1)/[\alpha + n_0(\mu - \mu_0)^2], \alpha + 1 \right\}.$$

The intuition motivating (3) is that since $\pi(y|\mu)$ has a heavier tail than $\pi(y|\theta)$, averages of reciprocal ordinates become averages of less variable quantities than in (2). Measuring stability as above, we observe that

$$\text{E} \left\{ [\pi(y|\mu)]^{-2} \middle| y \right\} \propto \int \frac{\{1 + [(y - \mu)^2 + n_0(\mu - \mu_0)^2]/\alpha\}^{\alpha/2+1}}{\{1 + n_0(\mu - \mu_0)^2/\alpha\}^{\alpha+1}} d\mu \quad (4)$$

is finite when $\alpha > 1$ and $n_0 > 0$. This result is proved in Appendix II.

Figure 1 compares the harmonic mean $\hat{\pi}_{\text{HM}}(y)$ to the stabilized harmonic mean $\hat{\pi}_{\text{SHM}}(y)$ for various parameter settings of this simple normal example. For each case, both estimates use a common sample of $B = 5,000$ independent and identically distributed posterior draws for the mean μ and precision ψ . Shown for each sample is the value of both estimators using ever larger amounts of the sample. Figure 1 shows clearly how the infinite variance of the harmonic mean estimator manifests itself in practice. Every so often a parameter value with a very small likelihood is generated from the posterior, and this yields a very large value of the reciprocal of the likelihood, which in turn greatly reduces $\hat{\pi}_{\text{HM}}(y)$. Subsequently, $\hat{\pi}_{\text{HM}}(y)$ increases gradually, until another very small likelihood is encountered. Improved performance of the stabilized harmonic mean is evident in Figure 1. The t -based estimator $\hat{\pi}_{\text{SHM}}(y)$ converges much more rapidly than the standard estimator, and does not exhibit the same pattern of occasional massive changes. To further validate this observation, we recomputed both final estimators on 100 independent posterior samples of size $B = 5000$ (Figure 2). Relative stability of the $\hat{\pi}_{\text{SHM}}(y)$ is clearly indicated.

The multivariate normal model is a direct extension of the univariate normal example discussed above. The standard estimator, obtained using equation (2), is a harmonic mean of multivariate normal densities. This can be easily shown to be an unstable estimator of the prior predictive density. Integrating the precision parameter leads to a heavier tailed multivariate t density, which can be used to obtain a stable estimator analogous to equation (3).

The stabilized harmonic mean was first reported in a statistical genetics application in which numerical stability of a t -based harmonic mean was observed (Satagopan, Yandell, Newton, and Osborn 1996). §3 presents a detailed study of this case. Although the genetic model used by these authors is rather specialized, the method to obtain a more stable estimate is quite general: approximate $\pi(y)$ by a harmonic mean of values $\pi[y|h(\theta^t)]$ where $\theta^1, \theta^2, \dots, \theta^B$ form a sample from the posterior distribution $\pi(\theta|y)$. The function $h(\theta)$ must reduce the parameter space as much as possible, while not making the calculation of the marginal likelihood $\pi[y|h(\theta)]$ too difficult. In the examples we work out, $h(\theta)$ is of lower dimension than θ , typically obtained by integrating out one or several of the components. Taking $h(\theta)$ to be constant is an extreme case; $\pi[y|h(\theta)]$ then becomes the integrated likelihood $\pi(y)$. Of course, if this were computable there would be no need to calculate an approximation, and in any case, the harmonic mean estimator would have zero variance. To form harmonic means from reduced distributions is a general variance reduction technique.

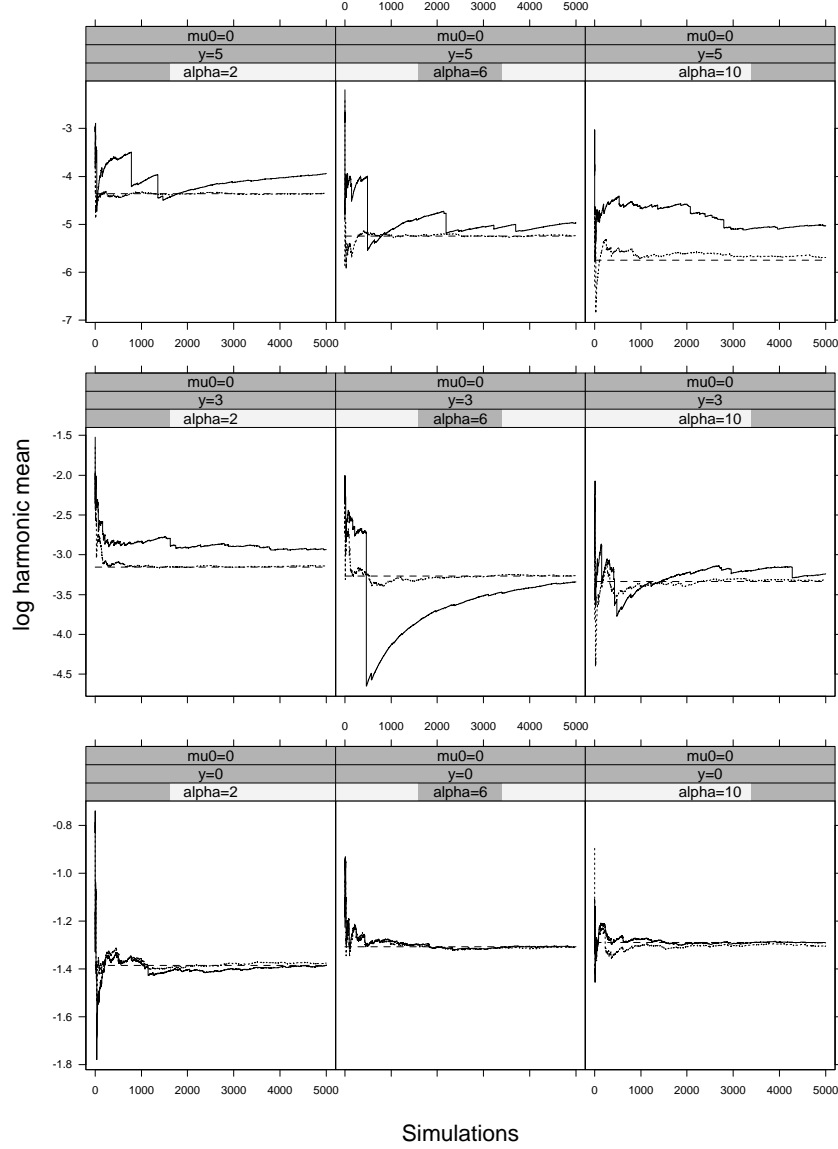


Figure 1: Logarithm of normal (bold line) and t -based (dotted line) harmonic mean estimates compared with the true value (dashed line), when the data y follow a univariate normal distribution as discussed in §1. The top row of the figure displays the harmonic mean estimates when $y = 5$ and $\mu_0 = 0$. The second row corresponds to $y = 0$ and $\mu_0 = 5$. The bottom row gives the figures for $y = 0$ and $\mu_0 = 0$. The 3 columns correspond to α values of 2, 6 and 10. The value of n_0 is 1.

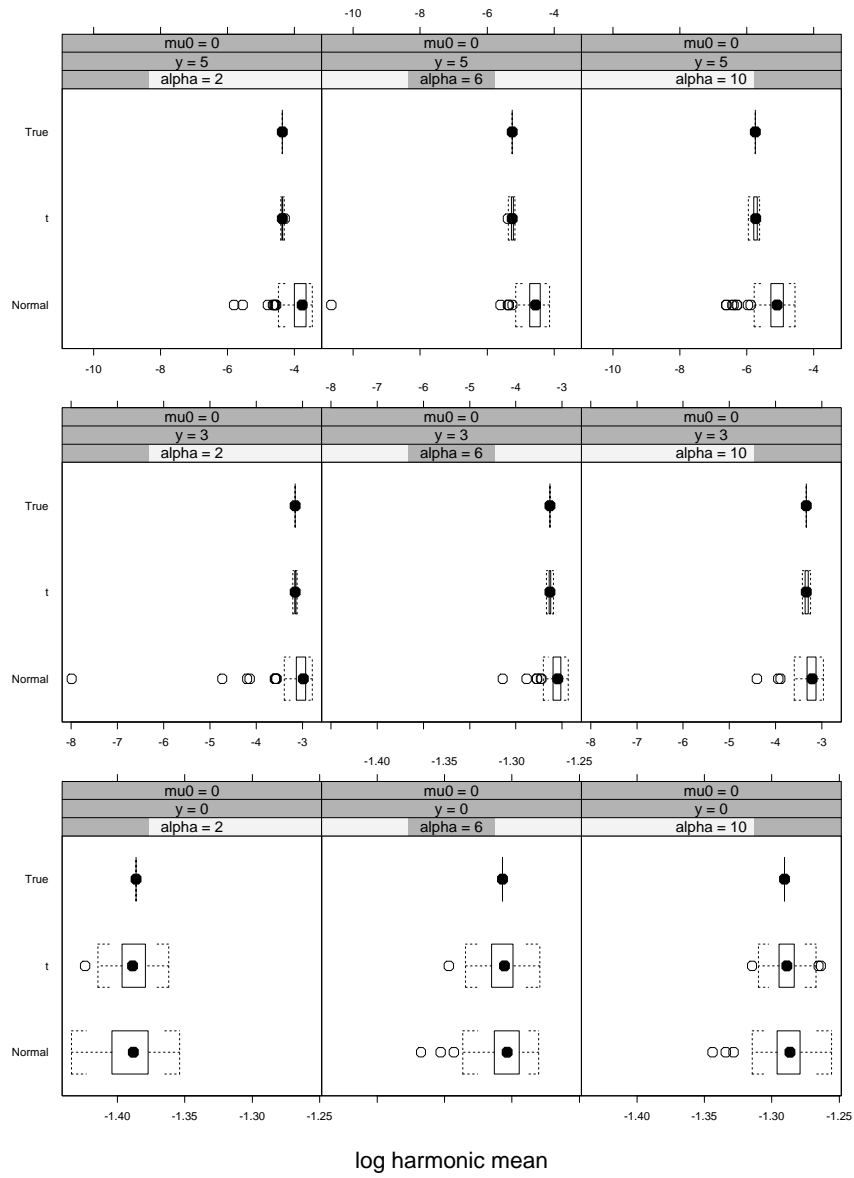


Figure 2: Boxplots to assess the variability of the estimated integrated likelihood. Shown are the true integrated likelihood, and the normal and t-based harmonic mean estimators on the logarithmic scale. The estimates are obtained from 100 Monte Carlo samples of size 5000. These estimates are shown for various configurations of parameters as in Figure 1.

Theorem 1 *If h is a measurable function of θ then*

$$\text{Var} \left\{ \frac{1}{\pi[y|h(\theta)]} \middle| y \right\} \leq \text{Var} \left\{ \frac{1}{\pi[y|\theta]} \middle| y \right\}.$$

Either variance may be infinite. If the left hand side is infinite, then the right hand side is infinite also.

To avoid measure-theoretic considerations, we prove Theorem 1 only under the additional condition that $h(\theta)$ is a dimension-reducing transformation: i.e. $\theta = (\alpha, \beta)$, $h(\theta) = \alpha$, and both α and β range freely so that the prior density $\pi(\theta) = \pi(\alpha)\pi(\beta|\alpha)$ is well-defined. See Appendix III for a proof. In certain hierarchical models, where analytical integration is possible on one or two levels, it may be possible to identify useful reductions $h(\theta)$ to facilitate stable harmonic mean calculations.

Gelfand and Dey (1994) noted an extension of the basic identity (1) which justifies estimating the integrated likelihood by the harmonic mean of $\pi(y|\theta^t)\pi(\theta^t)/f(\theta^t)$ where, as before, the θ^t 's are sampled from the posterior, but now $\pi(\theta)$ is the prior density and $f(\theta)$ is any (normalized) density on the parameter space. The idea is to choose f carefully so as to minimize Monte Carlo error. We show in §4 that our proposed stabilization can be combined with this technique for improved performance. Indeed there is some synergy in this combination because the proposed stabilization reduces the dimension of θ , thus making it simpler to identify a useful f function.

3 Statistical Genetics

Linear models are used frequently in quantitative genetics to relate variation in a measured trait (phenotype) to variation in underlying genes affecting the trait (genotype); Doerge, Zeng, and Weir (1997), for example, is a useful review from a statistics perspective. We reconsider the particular model

$$y_i = \mu + \sum_{j=1}^s \alpha_j g_{i,j} + \epsilon_i, \quad i = 1, \dots, n, \quad (5)$$

used by Satagopan *et al.* (1996) to infer the genetic causes of variation in the time-to-flowering phenotype in the plant species *Brassica napus*. In (5), the i indicates different plants in a sample of size $n = 105$, the phenotypes $y = (y_i)$ are the logarithms of the times to flowering, and the decomposition on the right-hand-side characterizes the expected phenotype conditional on the genotype $g_i = (g_{i,j})$ at a set of s different genetic loci. Here ϵ_i

is modeled as a mean zero normally distributed disturbance with variance σ^2 independent of genetic factors, μ is the marginal expected phenotype and α_j is the genetic effect of the j th quantitative trait locus (QTL). From the particular experimental design, each genotype $g_{i,j}$ takes one of two possible values, coded as $\{-1, 1\}$, with equal marginal probability.

The model (5) would be rather standard except that the genotypes $g = (g_i)$ are unobserved; in fact, for each i they represent the values of a random process defined over the whole genome and evaluated at s distinct positions $\lambda = (\lambda_1, \dots, \lambda_s)$, the s putative QTLs. The number of QTLs, s , is unknown, as are their positions λ and their effects $\alpha = (\alpha_1, \dots, \alpha_s)$. Indirect information about the QTL genotypes comes through genotype data $m = (m_i)$ obtained, in this example, from a panel of 10 molecular markers in the chromosomal region of interest. The statistical problem is to infer the unknown parameters $\theta = (\mu, \alpha, \lambda, \sigma^2)$ from marker and phenotype data (m, y) , and considering missing genotypes g .

Satagopan *et al.* (1996) presented a Bayesian solution in which Markov chain Monte Carlo (MCMC) was used to sample the posterior distribution of all the unknowns conditional on s , the number of QTLs, separately for a range of values of s . To infer s , the integrated likelihood $\pi(y|m, s)$ was approximated for each s via a harmonic mean, and this enabled calculation of Bayes factors

$$\text{BF}(s_1, s_2) = \pi(y|m, s_1)/\pi(y|m, s_2). \quad (6)$$

We reconsider this calculation in further detail. (Note that we can condition on marker information m because its marginal distribution $\pi(m)$ is not dependent on any of the unknown parameters.)

The prior for θ factorizes into a uniform prior over ordered loci $\lambda = (\lambda_1, \dots, \lambda_s)$ within the chromosomal region under consideration and a conjugate prior for μ , $\alpha = (\alpha_j)$, and σ^2 :

$$\begin{aligned} \pi(\mu|\sigma^2) &= \text{Normal}(\mu_0, \sigma^2/n_0), \\ \pi(\alpha_j|\sigma^2) &= \text{Normal}(\alpha_{0,j}, \sigma^2/n_{0,j}), \quad j = 1, \dots, s \\ \pi(\sigma^2) &= \text{Inverse Gamma}(\zeta/2, \zeta/2), \end{aligned}$$

where $\mu_0 = 5$, $n_0 = 1$, $\alpha_{0,j} = 5$, $n_{0,j} = 1$, for each j and $\zeta = 8$. Fixing the number of loci s , one complete scan of the MCMC sampler updates each element of θ and all the missing genotypes in g . See Satagopan *et al.* (1996) for further details on the component updates. A total of 3 chains, corresponding to $s = 1, 2, \text{ and } 3$, were obtained. For a fixed s ($= 1, 2, \text{ or } 3$), we report results below based on a chain of length 4000×100 complete scans, subsampled

every 100 scans, with the first 100 saved states removed as burn-in. This corresponds to an effective independent sample size of about 3900 for estimating the genetic effect parameters.

Unknowns (θ^t, g^t) are sampled from their posterior distribution conditional on observed phenotypes y , marker genotypes m and the model dimension parameter s . Invoking the standard harmonic mean argument, as in (2), we approximate $\pi(y|m, s)$ by

$$\hat{\pi}_{\text{HM}}(y|m, s) = \left[\frac{1}{B} \sum_{t=1}^B \frac{1}{\pi(y|m, \theta^t, g^t, s)} \right]^{-1}. \quad (7)$$

As in the simple normal example of §2, a problem arises with (7) because we are averaging reciprocals of normal ordinates. To stabilize the estimator, we integrate out the variance parameter σ^2 :

$$\hat{\pi}_{\text{SHM}}(y|m, s) = \left[\frac{1}{B} \sum_{t=1}^B \frac{1}{\pi(y|m, h(\theta^t), g^t, s)} \right]^{-1}, \quad (8)$$

where $h()$ returns all components of θ except the variance parameter. In (8), $\pi(y|m, h(\theta^t), g^t, s)$ is a scaled t density, $\text{St}_n(y|\mu + \alpha'g, I, \zeta)$.

Figure 3 shows the cumulative Bayes factor estimates obtained from three chains, ($s = 1, 2$, and 3), based on integrated likelihood estimates in either (7) and (8). Evidently the stabilization has worked in this more complicated example; there are fewer massive changes in the estimate. Numerically, we obtain $BF(1, 2) = 0.368$ using (7), and $BF(1, 2) = 0.395$ using the stabilized estimator (8). The estimates of $BF(2, 3)$ are rather more disparate: 13.15 and 4.39, respectively. In any case we would conclude that the two locus model is most likely *a posteriori*.

Figure 4 indicates the Monte Carlo sampling variability of the two estimators. The above computations were done on 75 repeated runs. To reduce the computational burden of the simulation, we used a value of B just half of the earlier value. The side-by-side boxplots further confirm the success of the stabilization in the present example.

We note that other dimension-reducing transformations $h(\cdot)$ could be used in this example. For example, we could sum out the genotype values g and thus average reciprocals of finite mixtures of normals (or t 's). It may also be possible to integrate the genetic effects α . Neither of these has been attempted here.

4 Beta–Binomial

A naturally occurring hierarchical model has observable counts $y = (y_i)$, $i = 1, \dots, m$, arising as conditionally independent binomial random variables with numbers of trials (n_i) and

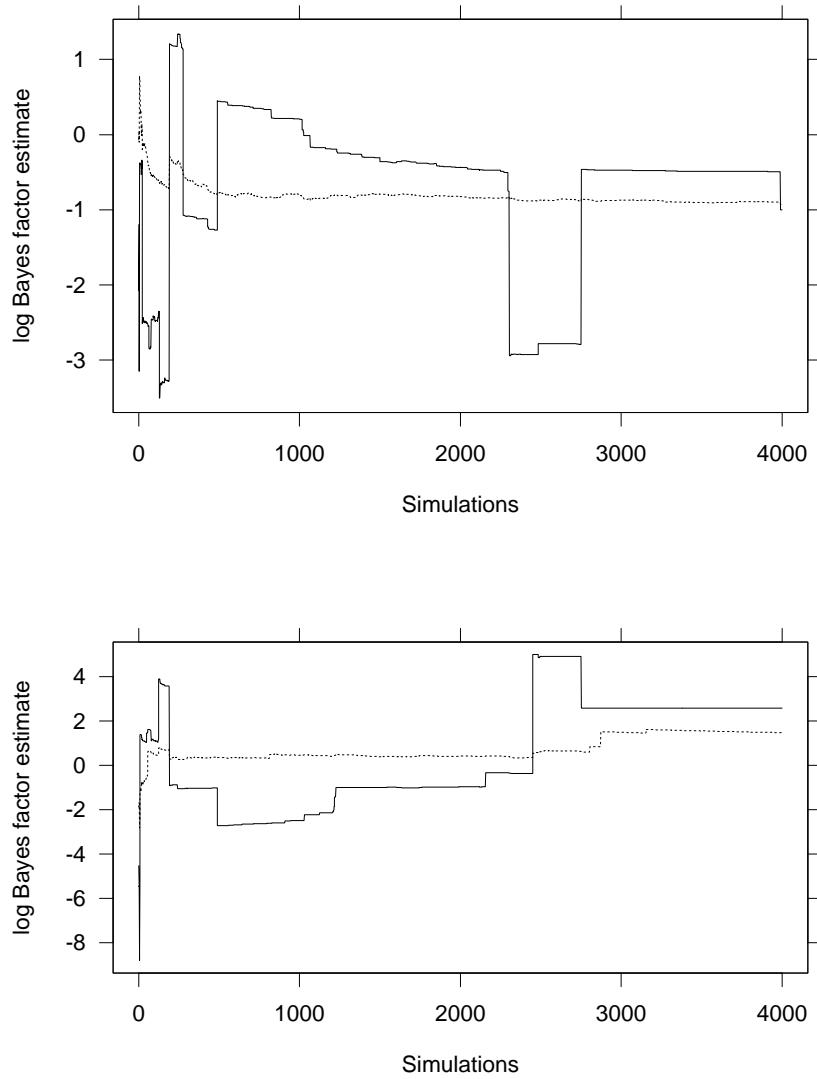


Figure 3: Bayes factors for the flowering time data discussed in §2. The comparison between one locus and two loci models is shown on the top. The bottom figure corresponds to the comparison between the two and three loci models. The bold line is the standard Bayes factor estimate. The dotted line is the t -based estimate.

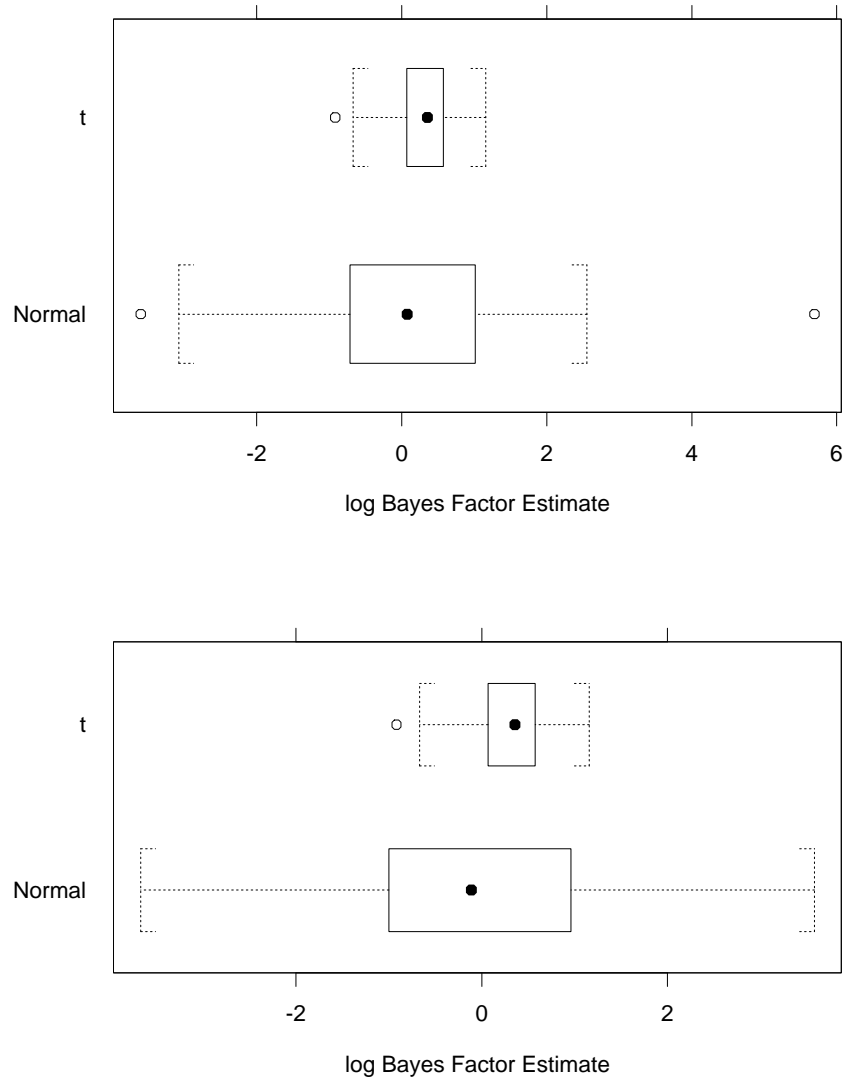


Figure 4: Assessing the variability of the Bayes factor estimates for the flowering time data using 75 repeated chains. Comparison between the one and two loci models are shown on the top, and comparison between two and three loci models are shown below. In each figure, variability among t-based estimates is shown in the top and that among the standard estimates is shown in the bottom.

success probabilities $p = (p_i)$. In turn, these (p_i) are modeled as conditionally independent beta variables with canonical hyperparameters a and b say, upon which some further prior distribution $\pi(a, b)$ is placed. To obtain the probability of y in this model, we must integrate out both (p_i) and the hyperparameters a and b . It is routine to sample the full parameter set $\theta = (p, a, b)$ from its posterior distribution (Gelman, Carlin, Stern, and Rubin 1996). For example, an MCMC simulation might update each p_i from its Beta full-conditional distribution, and then resort, perhaps, to a random-walk proposal to update a and b .

The basic harmonic mean combines reciprocals of binomial likelihoods from the posterior sample, and, it turns out, can be quite unstable. As before, stability is determined by the second noncentral moment

$$E \left\{ [\pi(y|\theta)]^{-2} | y \right\} \propto \int \int \prod_i \left\{ \int p^{a-1-y_i} (1-p)^{b-1-n_i+y_i} dp \right\} \pi(a, b) da db.$$

Unless we take an extreme prior $\pi(a, b)$ which ensures $a > \max(y_i)$ and $b > \max(n_i - y_i)$, this integral can diverge. The resulting prior is unrealistically peaked, which is unsatisfactory, ruling out the standard (unstabilized) harmonic mean estimator as a practical tool for the beta-binomial model.

It is straightforward to stabilize the harmonic mean by reducing the dimension of θ as in previous examples. One possibility is to take $h(\theta) = (a, b)$; i.e. to integrate out all the binomial success probabilities. In this conjugate structure, we have a closed form Beta-binomial expression for $\pi\{y|h(\theta)\}$, namely

$$\pi\{y|h(\theta)\} = \prod_i \frac{\Gamma(n_i + 1)}{\Gamma(n_i - y_i + 1)\Gamma(y_i + 1)} \frac{\Gamma(a + b)}{\Gamma(a + b + n_i)} \frac{\Gamma(a + y_i)}{\Gamma(a)} \frac{\Gamma(b + n_i - y_i)}{\Gamma(b)}. \quad (9)$$

The harmonic mean of these beta-binomial probabilities, calculated from the (a, b) 's sampled from their posterior, is consistent for the integrated likelihood. We may expect this to be more stable since the beta-binomial distribution is more diffuse than the binomial, and so the reciprocals of the probabilities may not be as extreme. The stability of this estimator is determined by the second noncentral moment, which satisfies

$$E \left\{ [\pi(y|a, b)]^{-2} | y \right\} \leq \int (a + b + n_{\max} - 1)^m \pi(a, b) da db,$$

where $n_{\max} = \max n_i$. Stability is ensured when prior moments of a and b exist.

Data on free-throw percentages from the National Basketball Association (NBA) provide an interesting demonstration of the harmonic mean calculations. On March 9, 1999, there were 414 active NBA players of whom 374 had attempted at least one free throw by that

point in the season. Among these 374 players, the numbers of attempts (n_i) ranged from 1 to 205, with a mean of about 35. We model y_i , the number of made free throws by player i , to be Binomial with n_i trials and unknown success probability p_i . The average free throw percentage y_i/n_i is about 70% in the data reported at www.yahoo.com (and available from the authors).

We consider the problem of evaluating the integrated likelihood $\pi(y)$ under the hierarchical beta-binomial model given above. This would be useful when comparing this model with other hypothesized models for these data. We place independent standard exponential priors on $a - \epsilon$ and $b - \epsilon$ where $\epsilon = 1$ is a lower truncation point of the prior. MCMC was used to simulate the posterior. The following numerical results are based on a single chain of length 50000×50 complete scans, subsampled every 50 scans, and with the first 100 saved states removed as burn-in. Significant trends were not detected in the output and time-series diagnostics indicated that little dependence remained in the saved states. Computations were done separately on a second run and we saw no appreciable differences (data not shown).

Natural logarithms of the product binomial likelihood and the product beta-binomial likelihood (9) are monitored in Figures 5a and 5b. From these values we obtain the standard harmonic mean estimate and the stabilized one. The log estimates are -820.8 and -944.2 respectively; these are quite different. The standard estimate is known to be unstable. Indeed the variance of the sampled loglikelihood values is 143.8 while that of the sampled log beta-binomial values is only 4.0. Variance on the log scale does not tell the whole story because we are averaging on the anti-log scale; it is outliers (having very low likelihood) that are particularly influential, but still variance gives some indication.

Suspecting that some additional improvements could be made, we combined the stabilization technique with the method discussed at the end of §2, using a Gaussian approximation to the posterior $\pi(a, b|y)$ as the density f . The estimate becomes a harmonic mean of the values $\pi(y|a, b)\pi(a, b)/f(a, b)$, with (a, b) 's sampled from their posterior. Figure 5c shows the time series of adjusted marginal likelihoods. The main advantage of this adjustment is that now the influence of individual sample points is greatly diminished. The estimated log integrated likelihood is -952.6. A brute force grid-based numerical integration of $\pi(y|a, b)\pi(a, b)$ gave -951.4. Thus we see that the initial stabilization method worked fairly well and was easily improved.

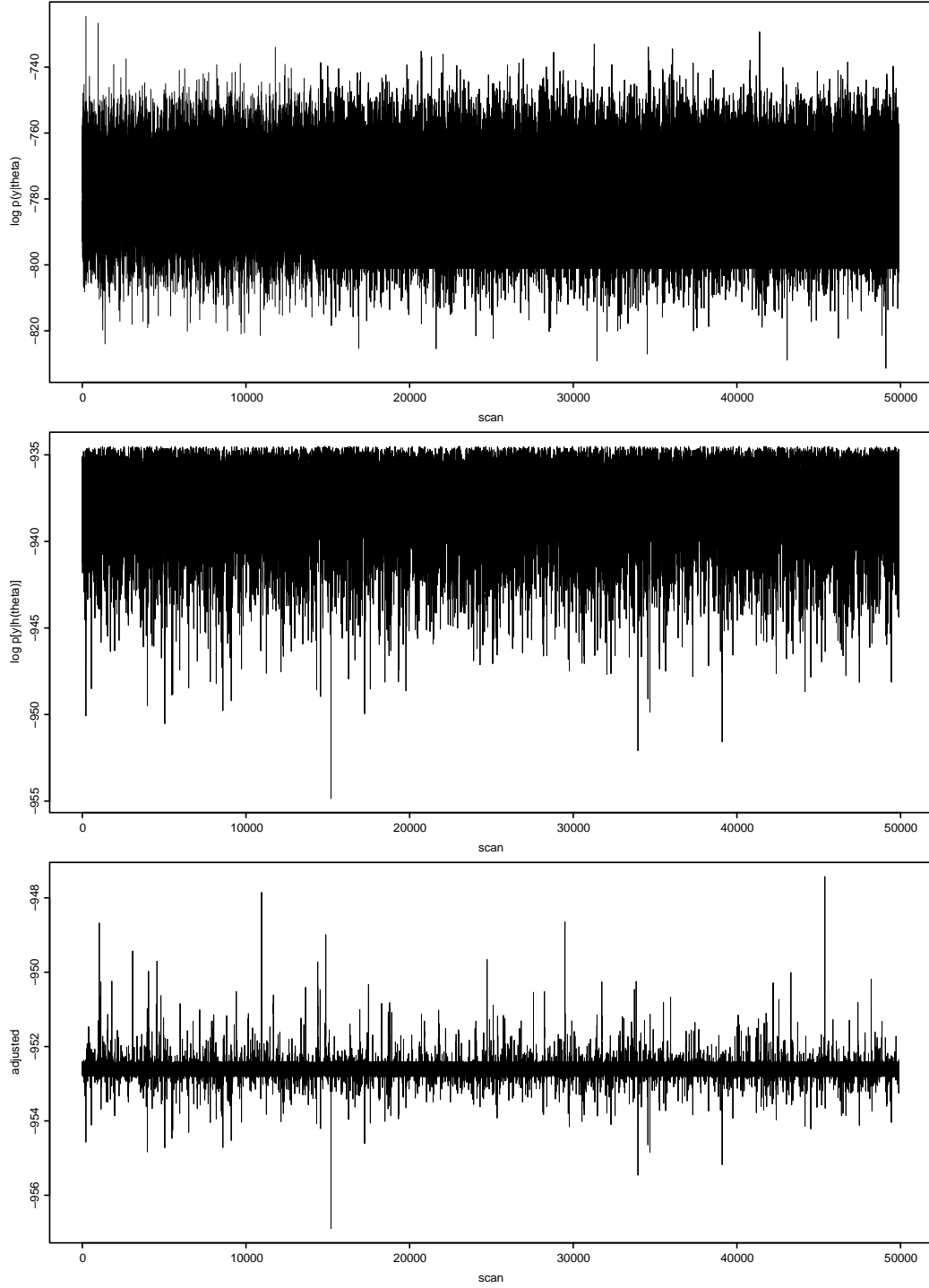


Figure 5: The top panel, middle panel and bottom panel compare the standard harmonic mean estimator, modified estimator, and adjusted estimator based on Gelfand and Dey (1994), respectively, for the NBA example.

5 Other Reductions: A Simple Poisson-Gamma Model

Sometimes useful reductions are hard to find, and the natural approach we have considered of integrating out a parameter does not work. A simple example is when y has a Poisson distribution with mean $\gamma\lambda$, and γ is exponentially distributed with mean 1 and independent of λ *a priori*. The standard harmonic mean estimator of $\pi(y)$ uses samples $\theta^i = (\lambda^i, \gamma^i)$ from $\pi(\theta|y)$, and averages the reciprocals of Poisson probabilities. Stability depends on the second noncentral moment

$$E \left\{ [\pi(y|\theta)]^{-2} \middle| y \right\} \propto \int \int \frac{1}{(\gamma\lambda)^y} \exp\{-\gamma(1-\lambda)\} \pi(\lambda) d\gamma d\lambda.$$

Note that the inner integral diverges for any $\lambda > 1$, so that the standard harmonic mean is unstable. The natural reduction would be to take $h(\theta) = \lambda$. Thus the marginal likelihood $\pi[y|h(\theta)] = \pi(y|\lambda)$ is a geometric distribution $\lambda^y/(1+\lambda)^{(y+1)}$. Stability here hinges upon

$$E \left\{ [\pi(y|\lambda)]^{-2} \middle| y \right\} \propto \int \left(\frac{1+\lambda}{\lambda} \right)^y (1+\lambda) \pi(\lambda) d\lambda.$$

For small λ , the dominant term of the integrand is $\pi(\lambda)/\lambda^y$, and so stability of the modified harmonic mean depends on the prior, though for a standard Gamma prior, for example, this integral can diverge. In other words, both variances in Theorem 1 equal infinity. Thus integrating out γ does not produce a stabilized harmonic mean estimator in this case.

Another, further reduction does work, however. Consider the case where λ , like γ , has a prior exponential distribution with mean 1. Suppose that $h(\theta) = 0$ if $\lambda \leq \epsilon$, and $h(\theta) = \lambda$ if $\lambda > \epsilon$, where ϵ is a small predetermined constant. Then $\pi[y|h(\theta) = 0] \approx \epsilon^{y+1}/(y+1)$ (better approximations are readily available if necessary), and it is easily shown that $E\{\pi[y|h(\theta)]^{-2}|y\} < \infty$. Thus, with this refinement, the modified harmonic mean estimator is stable.

6 Discussion

In this article we have described a way to stabilize the harmonic mean estimator of the integrated likelihood (Newton and Raftery 1994) by taking advantage of dimension-reducing transformations on the parameter space. The proposed variance stabilizing method extends a very simple tool into a range of widely used hierarchical statistical models. As illustrated in §3 and §4, dimension reduction is straight forward in certain hierarchical models. Sometimes the natural approach of integrating out a nuisance parameter does not yield a stabilized

estimator, however, and one must search farther. We have given one example, a simple Poisson-Gamma model, in §5 where the natural approach does not work directly, but a slight refinement of the $h(\cdot)$ function does yield a stabilized estimator. The trick used there to find this refined h function was based on the fact that the estimator is stable if and only if $E\{\pi[y|h(\theta)]^{-2}|y\} < \infty$. We wrote this expectation as an integral, identified the part of the range of integration responsible for the integral being infinite, and effectively carried out the integration over that small part of the space via analytic approximation, thus defining a new h function. Dimension reduction for variance stabilization may not be an effective method to compute normalizing constants in certain very hard problems. In the cases we have studied, we have shown that it is possible to stabilize the harmonic mean estimator and obtain estimates that are much more accurate, but still easy to calculate.

Gelfand and Dey (1994) derived an alternative approach to estimating the integrated likelihood $\pi(y)$, as discussed towards the end of §2. However, their estimator can be sensitive to the choice of the function $f(\cdot)$, and can also suffer from instability for the same reasons as the standard harmonic mean estimator does. In §4 we showed that their approach can be combined with the stabilization method proposed here for improved performance of the harmonic mean estimator.

Another application of the proposed stabilization approach includes robust linear models (Andrews and Mallows 1974, Carlin and Louis 1996). The robust linear model has an error term distributed as Z/\sqrt{U} , where Z and U are independently distributed as Normal(0, precision = ψ) and χ^2 with ν degrees of freedom, respectively. The standard harmonic mean estimator can have infinite variance. A stabilized harmonic mean estimator can be obtained by integrating out the denominator U (details not shown but available from the authors).

Hierarchical models which involve standard distributions may be good candidates for the present approach. For one thing, MCMC is well understood for within-model posterior simulation. Furthermore, the integrations required for dimension reduction may be solved analytically. The simplicity of the resulting stabilized harmonic mean is its main advantage.

The reversible-jump MCMC (Green 1995) is a specialized algorithm for Bayesian model selection. Satagopan and Yandell (1996) used this method to estimate the number of QTLs (s) in the statistical genetics problem of §3 by approximating the posterior model probabilities using the reversible-jump MCMC algorithm. The reversible jump MCMC algorithm involves additional calculations beyond single-model MCMC, and the resulting algebra and computation can be burdensome. Further, careful implementation of the algorithm includes an appropriate choice of prior distribution for s , as this can influence rapid mixing of the

chains (Satagopan and Yandell 1996). When the number of models considered is not too large, as for example in our QTL example, the methods described here enable one to infer the Bayes factors directly from single-model MCMC runs, which can be easier to implement than reversible jump MCMC.

Appendix I: Student's t

Student t

Copying Bernardo and Smith (1994, page 122),

$$\text{St}(x|\mu, \lambda, \alpha) = c \left[1 + \frac{\lambda}{\alpha} (x - \mu)^2 \right]^{-(\alpha+1)/2}$$

where

$$c = \frac{\Gamma((\alpha+1)/2)}{\Gamma(\alpha/2) \Gamma(1/2)} \left(\frac{\lambda}{\alpha} \right)^{1/2}.$$

Multivariate Student t

Using the notation of Bernardo and Smith (1994, page 139),

$$\text{St}_n(x|\mu, \lambda, \alpha) = c \left[1 + \frac{1}{\alpha} (x - \mu)^T \lambda (x - \mu) \right]^{-(\alpha+n)/2},$$

where

$$c = \frac{\Gamma((\alpha+n)/2)}{\Gamma(\alpha/2) (\alpha\pi)^{n/2}} \det(\lambda)^{1/2}.$$

x and μ are of dimension n . λ is a symmetric, positive-definite $n \times n$ matrix, and $\alpha > 0$.

Appendix II: Proof of equation 4

Define

$$f(\mu) = \frac{n_0}{\alpha} (\mu - \mu_0)^2 \quad \text{and} \quad g(\mu) = \frac{1}{\alpha} (y - \mu)^2.$$

Set

$$a(\mu) = 1 + \frac{g(\mu)}{1 + f(\mu)}.$$

It can be easily shown that the maxima of the continuous function $a(\mu)$ occurs at $\mu^* = \mu_0 - \alpha/[n_0(y - \mu_0)]$, and the maximum value of the function is

$$a(\mu^*) = 1 + \frac{1}{n_0} + g(\mu_0) .$$

Further $a(\mu) \rightarrow 1 + 1/n_0$, as $\mu \rightarrow \pm\infty$. The expected value of interest can be written as

$$E \left\{ \frac{1}{[\pi(y|\mu)]^2} y \right\} \propto \int [a(\mu)]^{\alpha/2+1} [1 + f(\mu)]^{-\alpha/2} d\mu ,$$

where $[1 + f(\mu)]^{-\alpha/2}$ is proportional to a t -density of the form

$$\text{St}(\mu|\mu_0, n_0(\alpha - 1)/\alpha, \alpha - 1) .$$

Since $1 \leq a(\mu) \leq a(\mu^*)$, the integral on the right hand side is finite by dominated convergence theorem when $\alpha > 1$ and $n_0 > 0$.

Appendix III: Proof of Theorem 1

Define $\alpha = h(\theta)$, write $\theta = (\alpha, \beta)$, and set

$$a = E \left\{ \frac{1}{[\pi(y|\alpha)]^2} y \right\} \quad \text{and} \quad b = E \left\{ \frac{1}{[\pi(y|\theta)]^2} y \right\} .$$

Since both $1/\pi(y|\alpha)$ and $1/\pi(y|\theta)$ have common expectation $1/\pi(y)$, it suffices to show that $a \leq b$. Expanding b , we have

$$\begin{aligned} b &= \int \int \frac{1}{[\pi(y|\alpha, \beta)]^2} \pi(\alpha, \beta|y) d\beta d\alpha \\ &= \int \int \frac{1}{[\pi(y|\alpha, \beta)]^2} \pi(\beta|\alpha, y) p(\alpha|y) d\beta d\alpha \\ &= \int b(\alpha) \pi(\alpha|y) d\alpha \end{aligned}$$

where

$$b(\alpha) = \int \frac{1}{[\pi(y|\alpha, \beta)]^2} \pi(\beta|\alpha, y) d\beta .$$

By contrast,

$$a = \int a(\alpha) \pi(\alpha|y) d\alpha$$

where

$$a(\alpha) = \frac{1}{[\pi(y|\alpha)]^2}.$$

Therefore, it is sufficient to prove that $a(\alpha) \leq b(\alpha)$ for all α . Simplifying $b(\alpha)$, we have

$$\begin{aligned} b(\alpha) &= \int \frac{1}{[\pi(y|\alpha, \beta)]^2} \pi(\beta|\alpha, y) d\beta \\ &= \int \frac{1}{[\pi(y|\alpha, \beta)]^2} \frac{\pi(y|\alpha, \beta) \pi(\beta|\alpha) \pi(\alpha)}{\pi(y|\alpha) \pi(\alpha)} d\beta \\ &= \frac{1}{\pi(y|\alpha)} \int \frac{\pi(\beta|\alpha)}{\pi(y|\alpha, \beta)} d\beta. \end{aligned}$$

Cancelling one factor $1/\pi(y|\alpha)$, we have $a(\alpha) \leq b(\alpha)$ if

$$\frac{1}{\pi(y|\alpha)} \leq \int \frac{\pi(\beta|\alpha)}{\pi(y|\alpha, \beta)} d\beta.$$

This follows by Jensen's inequality using the distribution $\pi(\beta|\alpha)$. In the event that one or another of the integrals diverges, $a(\alpha) \leq b(\alpha)$ must continue to hold.

References

- Andrews, D. F. and C. L. Mallows (1974). Scale mixtures of normality. *Journal of the Royal Statistical Society, Series B* 36, 99–102.
- Carlin, B. P. and T. A. Louis (1996). *Bayes and Empirical Bayes Methods for Data Analysis*, Chapter 6, pp. 209–211. Chapman and Hall.
- Chib, S. (1995). Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association* 90, 1313–1321.
- DiCiccio, T. J., R. E. Kass, A. E. Raftery, and L. Wasserman (1997). Computing Bayes factors by combining simulation and asymptotic approximation. *Journal of the American Statistical Association* 92, 903–915.
- Doerge, R. W., Z.-B. Zeng, and B. S. Weir (1997). Statistical issues in the search for genes affecting quantitative traits in experimental populations. *Statistical Science* 12, 195–219.
- Gelfand, A. E. and D. K. Dey (1994). Bayesian model choice: asymptotics and exact calculations. *Journal of the Royal Statistical Society, Series B* 56, 501–514.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (1996). *Bayesian Data Analysis*. Chapman and Hall.

Green, P. J. (1995). Reversible Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82, 711–732.

Kass, R. E. and A. E. Raftery (1995). Bayes factors. *Journal of the American Statistical Association* 90, 773–795.

Lewis, S. M. and A. E. Raftery (1997). Estimating Bayes factors via posterior simulation with the Laplace-Metropolis estimator. *Journal of the American Statistical Association* 92, 648–655.

Newton, M. A. and A. E. Raftery (1994). Approximate Bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society B* 56, 3–48.

Raftery, A. E. (1996). Hypothesis testing and model selection. In W. R. Gilks, S. Richardson, and D. J. Spiegelhalter (Eds.), *Markov chain Monte Carlo in practice*, Chapter 10, pp. 163–188. Chapman and Hall.

Satagopan, J. M. and B. S. Yandell (1996). Estimating the number of quantitative trait loci via bayesian model determination. In *Proceedings of the Section on Biometrics*. Joint Statistical Meetings, Chicago.

Satagopan, J. M., B. S. Yandell, M. A. Newton, and T. C. Osborn (1996). A Bayesian approach to detect quantitative trait loci using Markov chain Monte Carlo. *Genetics* 144, 805–816.